

09/615 726

## A METHOD AND MECHANISM FOR THE CREATION, MAINTENANCE, AND COMPARISON OF SEMANTIC ABSTRACTS

### ABSTRACT

5 Codifying the “most prominent measurement points” of a document can be used to  
measure semantic distances given an area of study (e.g., white papers on some subject area).  
A semantic abstract is created for each document. The semantic abstract is a semantic  
measure of the subject or theme of the document providing a new and unique mechanism for  
characterizing content. The semantic abstract includes state vectors in the topological vector  
10 space, each state vector representing one lexeme or lexeme phrase about the document. The  
state vectors can be dominant phrase vectors in the topological vector space mapped from  
dominant phrases extracted from the document. The state vectors can also correspond to  
words in the document that are most significant to the document’s meaning (the state vectors  
are called dominant vectors in this case). One semantic abstract can be directly compared  
15 with another semantic abstract, resulting in a numeric semantic distance between the  
semantic abstracts being compared.

09/615 726